685810484

PATENT COOPERATION TREATY

PCT

INTERNATIONAL PRELIMINARY REPORT ON PATENTABILITY

(Chapter II of the Patent Cooperation Treaty)

(PCT Article 36 and Rule 70)

Applicant's or agent's file reference FP19456	**FOR FURTHER ACTION**	See Form PCT/IPEA/416
International application No. **PCT/AU2004/000696**	International filing date *(day/month/year)* 26 May 2004	Priority date *(day/month/year)* 26 May 2003

International Patent Classification (IPC) or national classification and IPC

Int. Cl. ^{7} G06F 17/18, G06F 19/00, G06F 159:00

Applicant

COMMONWEALTH SCIENTIFIC AND INDUSTRIAL RESEARCH ORGANISATION et al

1. This report is the international preliminary examination report, established by this International Preliminary Examining Authority under Article 35 and transmitted to the applicant according to Article 36.

2. This REPORT consists of a total of 3 sheets, including this cover sheet.

3. This report is also accompanied by ANNEXES, comprising:

   a. ☒ *(sent to the applicant and to the International Bureau)* a total of 9 sheets, as follows:

      ☒ sheets of the description, claims and/or drawings which have been amended and are the basis for this report and/or sheets containing rectifications authorized by this Authority (see Rule 70.16 and Section 607 of the Administrative Instructions).

      □ sheets which supersede earlier sheets, but which this Authority considers contain an amendment that goes beyond the disclosure in the international application as filed, as indicated in item 4 of Box No. I and the Supplemental Box.

   b. □ *(sent to the International Bureau only)* a total of (indicate type and number of electronic carrier(s)) , containing a sequence listing and/or table related thereto, in computer readable form only, as indicated in the Supplemental Box Relating to Sequence Listing (see Section 802 of the Administrative Instructions).

4. This report contains indications relating to the following items:

☒	Box No. I	Basis of the report
□	Box No. II	Priority
□	Box No. III	Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
□	Box No. IV	Lack of unity of invention
☒	Box No. V	Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement
□	Box No. VI	Certain documents cited
□	Box No. VII	Certain defects in the international application
□	Box No. VIII	Certain observations on the international application

Date of submission of the demand 30 November 2004	Date of completion of the report 26 August 2005
Name and mailing address of the IPEA/AU AUSTRALIAN PATENT OFFICE PO BOX 200, WODEN ACT 2606, AUSTRALIA E-mail address: pct@ipaustralia.gov.au Facsimile No. (02) 6285 3929	Authorized Officer **DALE SIVER** Telephone No. (02) 6283 2196

INTERNATIONAL PRELIMINARY REPORT ON PATENTABILITY

International application No.
PCT/AU2004/000696

Box No. I **Basis of the report**

1. With regard to the **language**, this report is based on the international application in the language in which it was filed, unless otherwise indicated under this item.

   □ This report is based on translations from the original language into the following language ,
   which is the language of a translation furnished for the purposes of:

   □ international search (under Rules 12.3 and 23.1 (b))

   □ publication of the international application (under Rule 12.4)

   □ international preliminary examination (under Rules 55.2 and/or 55.3)

2. With regard to the **elements** of the international application, this report is based on (*replacement sheets which have been furnished to the receiving Office in response to an invitation under Article 14 are referred to in this report as "originally filed" and are not annexed to this report*):

   □ the international application as originally filed/furnished

   ☒ the description:

   pages **1,2,4-18,20,22-110** as originally filed/furnished
   pages* **3,19,21** received by this Authority on **17 August 2005** with the letter of **16 August 2005**
   pages* received by this Authority on with the letter of

   ☒ the claims:

   pages **112** as originally filed/furnished
   pages* as amended (together with any statement) under Article 19
   pages* **113,114** received by this Authority on **29 March 2005** with the letter of **29 March 2005**
   pages* **111,115,116,117** received by this Authority on **17 August 2005** with the letter of **16 August 2005**

   ☒ the drawings:

   pages **1/6 to 6/6** as originally filed/furnished
   pages* received by this Authority on with the letter of
   pages* received by this Authority on with the letter of

   □ a sequence listing and/or any related table(s) - see Supplemental Box Relating to Sequence Listing.

3. □ The amendments have resulted in the cancellation of:

   □ the description, pages
   □ the claims, Nos.
   □ the drawings, sheets/figs
   □ the sequence listing (*specify*):
   □ any table(s) related to the sequence listing (*specify*):

4. □ This report has been established as if (some of) the amendments annexed to this report and listed below had not been made, since they have been considered to go beyond the disclosure as filed, as indicated in the Supplemental Box (Rule 70.2(c)).

   □ the description, pages
   □ the claims, Nos.
   □ the drawings, sheets/figs
   □ the sequence listing (*specify*):
   □ any table(s) related to the sequence listing (*specify*):

* *If item 4 applies, some or all of those sheets may be marked "superseded."*

INTERNATIONAL PRELIMINARY REPORT ON PATENTABILITY

International application No.
PCT/AU2004/000696

Box No. V Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

1. Statement

Novelty (N)	Claims **1-26**	**YES**
	Claims	**NO**
Inventive step (IS)	Claims **1-26**	**YES**
	Claims	**NO**
Industrial applicability (IA)	Claims **1-26**	**YES**
	Claims	**NO**

2. Citations and explanations (Rule 70.7)

D1 WO 2001/018667 A2 (MICROSOFT CORPORATION) 15 March 2001
2 WO 2003/034270 A1 (Commonwealth Scientific and Industrial Research Organisation) 24 April 2003
D3 WO 2002/087431 A1 (STRUCTURAL BIOINFORMATICS, INC. et al.) 7 November 2002
D4 FIGUEIREDO M.A.T. "Bayesian learning of sparse classifiers" 2001
D5 US 6059724 A (CAMPELL et al.) 9 May 2000

Novelty (N)

D1 discloses a Relevance Vector Machine that does not use Jeffrey's hyperprior. D1 discloses how training sets are used in data modelling. See page 1 lines 17-28 where there is explicit disclosure of a training set (training data set in the form of input vectors and output vectors). The model is built upon the training data set although not entirely (for the reasons given on page 1 lines 22-24).
The description of D1 contains numerous other explicit disclosures of how training data is used with either the Support Vector Machine (SVM) or the Relevance Vector Machine (RVM). Each of these learning machines is intended to have the values of the weights adjusted from the training data sets. On page 7 it is stated (lines 1-3) that the learning machine accepts a training set of data and outputs a posterior distribution p(c|x) which is used later as a basis for the weights (eg. in the Relevance Vector Regression). The training data is used to build the model from the posterior distribution as well as from the prior distribution. The RVM uses an “ARD Gaussian prior within the art over the weights”.

Amended claim 1 includes (inter alia) the limitation that the hyperprior is based on a combined Gaussian distribution and Gamma hyperprior. This features confers novelty onto the claim and other similarly amended claims.

D2 discloses similar methods including a model, hyperprior and training sets. These features are combined to identify diagnostic components of a system. The limitation “that the hyperprior is based on a combined Gaussian distribution and Gamma hyperprior” is not found in this prior art, hence the claims now satisfy PCT rules for novelty.

Inventive step (IS)

None of the citations or obvious combination of the above documents disclose the method and apparatus as currently claimed.

Industrial applicability (IA)

The claims have an industrial application (for example organising databases for response groups) biological data using existing methods is time consuming, prone to false results and requires large amounts of computer memory if a meaningful result is to be obtained from the data. This is problematic in large scale screening scenarios where rapid and accurate screening is required.

It is therefore desirable to have a method, in particular for analysis of biological data, and more generally, for an improved method of analysing data from a system in order to predict a feature of interest for a sample from the system.

SUMMARY OF THE INVENTION

According to a first aspect of the present invention, there is provided a method of identifying a subset of components of a system based on data obtained from the system using at least one training sample from the system, the method comprising the steps of:

obtaining a linear combination of components of the system and weightings of the linear combination of components, the weightings having values based on the data obtained from the system using the at least one training sample, the at least one training sample having a known feature;

obtaining a model of a probability distribution of the known feature, wherein the model is conditional on the linear combination of components;

obtaining a prior distribution for the weighting of the linear combination of the components, the prior distribution comprising a hyperprior having a high probability density close to zero, the hyperprior being such that it is based on a combined Gaussian distribution and Gamma hyperprior;

combining the prior distribution and the model to generate a posterior distribution; and

identifying the subset of components based on a set of the weightings that maximise the posterior distribution.

carbohydrates, lipids or any other measurable component of the subject.

In a particularly embodiment of the fifth aspect, the compound is a pharmaceutical compound or a composition comprising a pharmaceutical compound and a pharmaceutically acceptable carrier.

The identification method of the present invention may be implemented by appropriate computer software and hardware.

According to a sixth aspect of the present invention, there is provided an apparatus for identifying a subset of components of a system from data generated from the system from a plurality of samples from the system, the subset being capable of being used to predict a feature of a test sample, the apparatus comprising:

a processing means operable to:

obtain a linear combination of components of the system and obtain weightings of the linear combination of components, each of the weightings having a value based on data obtained from at least one training sample, the at least one training sample having a known feature;

obtaining a model of a probability distribution of a second feature, wherein the model is conditional on the linear combination of components;

obtaining a prior distribution for the weightings of the linear combination of the components, the prior distribution comprising an adjustable hyperprior which allows the prior probability mass close to zero to be varied wherein the hyperprior is based on a combined Gaussian distribution and Gamma hyperprior;

combining the prior distribution and the model to generate a posterior distribution; and

identifying the subset of components having component weights that maximize the posterior distribution.

on a computing device, allows the computing device to carry out a method of identifying components from a system that are capable of being used to predict a feature of a test sample from the system, and wherein a linear combination of components and component weights is generated from data generated from a plurality of training samples, each training sample having a known feature, and a posterior distribution is generated by combining a prior distribution for the component weights comprising an adjustable hyperprior which allows the probability mass close to zero to be varied wherein the hyperprior is based on a combined Gaussian distribution and Gamma hyperprior, and a model that is conditional on the linear combination, to estimate component weights which maximise the posterior distribution.

Where aspects of the present invention are implemented by way of a computing device, it will be appreciated that any appropriate computer hardware e.g. a PC or a mainframe or a networked computing infrastructure, may be used.

According to a twelfth aspect of the present invention, there is provided a method of identifying a subset of components of a biological system, the subset being capable of predicting a feature of a test sample from the biological system, the method comprising the steps of:

obtaining a linear combination of components of the system and weightings of the linear combination of components, each of the weightings having a value based on data obtained from at least one training sample, the at least one training sample having a known first feature;

obtaining a model of a probability distribution of a second feature, wherein the model is conditional on the linear combination of components;

obtaining a prior distribution for the weightings of the linear combination of the components, the prior distribution comprising an adjustable hyperprior which allows the probability mass close to zero to be varied;

THE CLAIMS DEFINING THE INVENTION ARE AS FOLLOWS:

1. A method of identifying a subset of components of a system based on data obtained from the system using at least one training sample from the system, the method comprising the steps of:

obtaining a linear combination of components of the system and weightings of the linear combination of components, the weightings having values based on data obtained from the at least one training sample, the at least one training sample having a known feature;

obtaining a model of a probability distribution of the known feature, wherein the model is conditional on the linear combination of components;

obtaining a prior distribution for the weighting of the linear combination of the components, the prior distribution comprising a hyperprior having a high probability density close to zero, the hyperprior being such that it is based on a combined Gaussian distribution and Gamma hyperprior;

combining the prior distribution and the model to generate a posterior distribution; and

identifying the subset of components based on a set of the weightings that maximise the posterior distribution.

2. The method as claimed in claim 1, wherein the step of obtaining the linear combination comprises the step of using a Bayesian statistical method to estimate the weightings.

3. The method as claimed in claim 1 or 2, further comprising the step of making an apriori assumption that a majority of the components are unlikely to be components that will form part of the subset of components.

4. The method as claimed in any one of the preceding claims, wherein the hyperprior comprises one or more adjustable parameters that enable the prior distribution near zero to be varied.

of:

$$l\left(\underset{\sim}{t} \mid \underset{\sim}{\beta}\right) = \prod_{j=l}^{N} \left(\frac{exp\left(Z_j \underset{\sim}{\beta}\right)}{\sum\limits_{i \in \Re_j} exp\left(Z_i \underset{\sim}{\beta}\right)} \right)^{d_j}$$

11. The method as claimed in claim 7, wherein the model based on the Parametric Survival model is in the form of:

$$L = \sum_{i=1}^{N} \left\{ c_i \, log\left(\mu_i\right) - \mu_i + c_i \left(log \left(\frac{\lambda\left(y_i\right)}{\Lambda\left(y_i ; \underset{\sim}{\varphi}\right)} \right) \right) \right\}$$

12. The method as claimed in any one of the preceding claims, wherein the step of identifying the subset of components comprises the step of using an iterative procedure such that the probability density of the posterior distribution is maximised.

13. The method as claimed in claim 12, wherein the iterative procedure is an EM algorithm.

14. A method for identifying a subset of components of a subject which are capable of classifying the subject into one of a plurality of predefined groups, wherein each group is defined by a response to a test treatment, the method comprising the steps of:

exposing a plurality of subjects to the test treatment and grouping the subjects into response groups based on responses to the treatment;

measuring components of the subjects; and

identifying a subset of components that is capable of classifying the subjects into response groups using the method as claimed in any one of claims 1 to 13.

15. An apparatus for identifying a subset of components of a subject, the subset being capable of being used to classify the subject into one of a plurality of predefined response groups wherein each response group, is formed by exposing a plurality of subjects to a test treatment and grouping the subjects into response groups based on the response to the treatment, the apparatus comprising:

an input for receiving measured components of the subjects; and

processing means operable to identify a subset of components that is capable of being used to classify the subjects into response groups using the method as claimed in any one of claims 1 to 13.

16. A method for identifying a subset of components of a subject that is capable of classifying the subject as being responsive or non-responsive to treatment with a test compound, the method comprising the steps of:

exposing a plurality of subjects to the test compound and grouping the subjects into response groups based on each subjects response to the test compound;

measuring components of the subjects; and

identifying a subset of components that is capable of being used to classify the subjects into response groups using the method as claimed in any one of claims 1 to 13.

17. An apparatus for identifying a subset of components of a subject, the subset being capable of being used to classify the subject into one of a plurality of predefined response groups wherein each response group is formed by exposing a plurality of subjects to a compound and grouping the subjects into response groups based on the response to the compound, the apparatus comprising;

an input operable to receive measured components of the subjects;

processing means operable to identify a subset of components that is capable of classifying the subjects into response groups using the method as claimed in any one of claims 1 to 13.

18. An apparatus for identifying a subset of components of a system from data generated from the system from a plurality of samples from the system, the subset being capable of being used to predict a feature of a test sample, the apparatus comprising:

a processing means operable to:

obtain a linear combination of components of the system and obtain weightings of the linear combination of components, each of the weightings having a value based on data obtained from at least one training sample, the at least one training sample having a known feature;

obtaining a model of a probability distribution of a second feature, wherein the model is conditional on the linear combination of components;

obtaining a prior distribution for the weightings of the linear combination of the components, the prior distribution comprising an adjustable hyperprior which allows the prior probability mass close to zero to be varied wherein the hyperprior is based on a combined Gaussian distribution and Gamma hyperprior;

combining the prior distribution and the model to generate a posterior distribution; and

identifying the subset of components having component weights that maximize the posterior distribution.

19. The apparatus as claimed in claim 18, wherein the processing means comprises a computer arranged to execute software.

20. A computer program which, when executed by a computing apparatus, allows the computing apparatus to carry out the method as claimed in any one of claims 1 to 13.

21. A computer readable medium comprising the computer program as claimed in claim 20.

22. A method of testing a sample from a system to identify a feature of the sample, the method comprising the steps of testing for a subset of components that are diagnostic of the feature, the subset of components having been determined by using the method as claimed in any one of claims 1 to 13.

23. The method as claimed in claim 22, wherein the system is a biological system.

24. An apparatus for testing a sample from a system to determine a feature of the sample, the apparatus comprising means for testing for components identified in accordance with the method as claimed in any one of claims 1 to 13.

25. A computer program which, when executed by on a computing device, allows the computing device to carry out a method of identifying components from a system that are capable of being used to predict a feature of a test sample from the system, and wherein a linear combination of components and component weights is generated from data generated from a plurality of training samples, each training sample having a known feature, and a posterior distribution is generated by combining a prior distribution for the component weights comprising an adjustable hyperprior which allows the probability mass close to zero to be varied wherein the hyperprior is based on a combined Gaussian distribution and Gamma hyperprior, and a model that is conditional on the linear combination, to estimate component weights which maximise the posterior distribution.

26. A method of identifying a subset of components of a biological system, the subset being capable of predicting a feature of a test sample from the biological system, the method comprising the steps of:

obtaining a linear combination of components of the system and weightings of the linear combination of components, each of the weightings having a value based on data obtained from at least one training sample, the at least one training sample having a known feature;

obtaining a model of a probability distribution of the known feature, wherein the model is conditional on the linear combination of components;

obtaining a prior distribution for the weightings of the linear combination of the components, the prior distribution comprising an adjustable hyperprior which allows the probability mass close to zero to be varied;

combining the prior distribution and the model to generate a posterior distribution; and

identifying the subset of components based on the weightings that maximize the posterior distribution.

DATED this 15th day of August 2005
CSIRO
By their Patent Attorneys
GRIFFITH HACK